Indian Health Service

National Data Warehouse

Data Integrity Verification Plan

Design Document

Version 1.0

July 2004

Department of Health and
Human Services

Indian Health Service

Information Technology
Support Center (ITSC)

Contents

Version Control

Version	Date	Notes
1.0	July 19, 2004	Initial Publish

Overview

The Indian Health Service National Data Warehouse (IHS NDW) is in the process of loading healthcare information from both government and commercial healthcare sources. This central database will provide a historical repository of patient registration and encounter information dating back to October 2000.

To ensure data integrity in the loading of data from multiple and varied operational systems to the NDW, ITSC has developed a Data Integrity Verification plan that is outlined in this document.

The IHS ITSC Data Integrity Verification Plan is comprised of the following:

Ensure data integrity
Verify that the data received and processed in the NDW accurately reflects the data transmitted from the source system. This includes multiple data checks as the data is processed into the NDW.

Ensure HIPAA compliance
Although HIPAA compliance is not primarily a data integrity issue, this plan will also ensure HIPAA requirements are met during the transmission and processing of patient-sensitive data.

Customer Reporting
Ensure the customer is kept apprised of any data issues that arise, as well as providing confirmation once data has been successfully processed and loaded into the NDW.

Test Strategy

The NDW test strategy utilizes the following levels of testing:

- ➢ Source to Target Counts

- ➢ Source to Target Data Verification

- ➢ Field to Field Verification

- ➢ Exception Processing

- ➢ Customer Reporting & Verification

Source to Target Counts

This process verifies that the number of records received from the source system matches the number of records received, and ultimately processed, into the data warehouse.

Source to Target Data Verification

This process verifies that the source field threshold is not subject to truncation during the transformation or loading of data.

Field to Field Verification

This process verifies the field values from the source system to target. This process ensures that the data mapping from the source system to the target is correct, and that data sent has been loaded accurately. **NOTE**: Source system data is accepted "as is"; i.e., the NDW relies on the facility to ensure the quality of the data that is being transmitted. The NDW verifies the integrity of the data. The NDW will, however, provide information to local sites that will assist them in improving the "quality" of their data.

Exception Processing

Because the NDW attempts to load as much customer data as possible, exception processing looks only for serious data errors that would cause system processing failures or data corruption. An Exception report verifying the number and types of errors encountered is produced and reviewed for additional processing and / or reporting to the customer.

Customer Reporting & Verification

After all initial data is loaded for an Area, a summary report will be generated and forwarded to the customer (the source system) for their review. By supplying this report, the customer is afforded a view to alert the NDW team to potential discrepancies.

Assumptions

Processing in RPMS Sites

- ➤ Initially, the NDW will begin loading all RPMS-based sites, utilizing the HL7 V2.4 protocol for data structure and transmission. Initial loads will consist of all Patient Registration data, and Encounter data dating from October 1, 2000. Incremental loads of data will be scheduled after initial loads have been successfully received and processed in the NDW.

- ➤ HL7 messages will be 'batched', and transmitted to the IHS Integration Engine (IE). The IE will 'translate' the HL7 messages into NDW-formatted flat ASCII files for further processing by the NDW Orientation Engine (OE).

- ➤ At the time of the RPMS extract, an Audit file will be produced. This file will be sent via FTP to a designated site, then used to do an automated "compare" against the extracted information (which will have been translated from HL7 messages into the NDW-formatted file). Any discrepancies will be noted and addressed as appropriate.

- ➤ The ETL will also perform particular data checks, ensuring data field sizes and types are accurate.

- ➤ In addition to the first automated compare, once the data has been loaded to the NDW, an automated compare will again be run, comparing the loaded data to the NDW-formatted file. Discrepancies will be noted and addressed as appropriate.

Processing in Non-RPMS Sites

Still under development.

Loading data from non-RPMS sites is still being reviewed. The QURE 2 process (described in this document, "QURE 2") will be utilized to compare data passed through the Integration Engine to the NDW. Whether a site will be able or willing to produce an

audit file may determine whether we will be able to verify source system data to NDW received data.

The "QURE" Process

The QURE (**Qu**ality **Re**conciliation) application is comprised of two (2) processes:

For RPMS sites only, **QURE 1** was developed to automatically compare incoming extracted data (converted to NDW-formatted flat ASCII files) against an Audit file created during the extract process (at the facility). This will ensure that the data transmissions and the Integration Engine processing did not alter data in an unintended manner. This program compares both Patient Registration and Patient Encounter fields, determines any discrepancies in the records, and then provides preliminary, internal use reports. These reports are then reviewed by the QA team, discrepancies are addressed as appropriate (e.g., 'bad' date formats, etc.), then logged. (See Appendix A. for an example of the Tracking Log data captured.) QURE 1 was developed to automatically compare Initial Load data (wherein Patient Registration and Patient Encounter files are sent in two separate files) as well as Incremental Load data (wherein Patient Registration and Patient Encounter files are sent in one file).

QURE 2 was developed to automatically compare the extracted data against the NDW load files, ensuring that the Orientation Engine and ETL processing did not skew data. QURE 2 generates internal use reports listing any data discrepancies. Discrepancies are addressed and corrected as required. The detail is then logged in the Tracking Log, and a final summary report is sent to the Facility (see Appendix B. for an example of the Final Summary Report).

Load Exception Reports are produced after the completion of the NDW load for each export. The load process writes exceptions to a database table, categorized as "Rejected", "Transformed", or "Informational".[1] These exceptions are then summarized and sent to the Facility for review. A sample Load Exception Report is displayed in Appendix C. A data mart is also being developed to give customers web-based access to exception report information.

---

[1] Examples of "rejected" records include encounter records linked to a registration record ID that does not exist in the NDW or records of chart numbers with no associated patient registration record. Examples of "transformed" records are records with 6-digit dates converted to 8-digit dates or with a code that exceeds the established field length and so is converted by the ETL program to a "null." An example of an "informational" record would be one with unusual values such as a registration record that does not have any associated chart numbers.
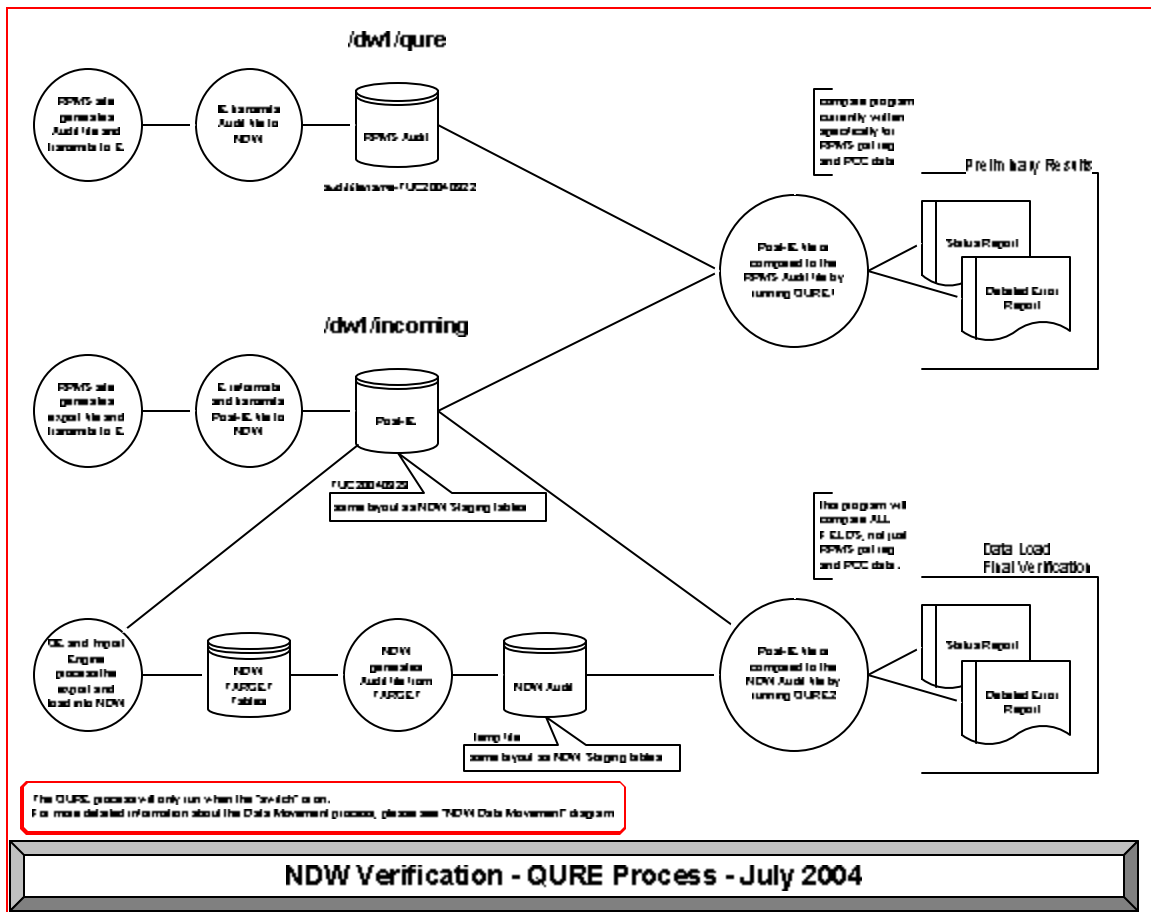
QURE Process Flow

Figure A.

Figure A. depicts the QURE process flow currently in place for both Patient Registration and Patient Encounter data loads. **NOTE**: QURE 1 is only used for RPMS-based facility extracts.

1. Extract and Audit files are sent to the IE from the Facility.

2. The IE processes and transmits both files to the NDW. The Extract file from the facility has been re-formatted, creating a "Post-IE" file.

3. The two (2) files are compared by QURE 1. Preliminary reports are generated (for internal use only) which are reviewed to determine whether the load process should continue, or whether a re-transmittal of the file is necessary. Appendix D. lists the data elements currently being compared for QURE 1.

4. If it is determined by review of the QURE 1 report that the load process should continue, the file is processed by the Orientation Engine (OE) and Import Engine, and then loaded to the NDW target tables.

5. The NDW generates a Data Integrity file; QURE 2 then compares the NDW file to the Post-IE file, creating reports (for internal use only) which are reviewed to determine whether the load process was successful.

   It should be noted that although not reflected in Figure A., there will be additional reports generated for the customer.

NDW Data Movement Process Flow

NDW Data Movement July 2004

3. The OE ACK (Orientation Engine Acknowledgement) process assigns an Export ID to the batch, logs information about the export, verifies the record counts, and e-mails acknowledgment to the sender. If errors are detected, the file is rejected and the sender is notified.

4. If the Post-IE file passes through the OE ACK process without error, the file is moved to the Pre-QURE directory. If the QURE is set to run automatically (i.e., the switch is "on"), QURE 1 is run at this time. The file will then be moved to the Preload directory (whether QURE is run or not).

5. The Post-IE file is then picked up by the Import Engine, which appends the Export ID to each record. If the Post-IE file contains encounter data, the Import Engine will then assign a unique Encounter Snapshot ID to each visit.

6. The Import Engine then performs specified data checks, including date formatting, data type errors, etc.

7. The Import Engine then updates the information that is logged for this export, reports any errors (if any), and loads the data into the NDW Target and Archive tables.

Appendix A. QURE Log / Data Tracking Elements

The QURE Tracking Log was developed to ensure A) export files were tracked from creation through load, B) files could be 'back-tracked' to the original export, if required, and C) time estimates could be provided to facilities to assist them in determining processing timelines.

The following factors are tracked:

- Area (e.g., Tucson)
- Facility (e.g., San Xavier)
- Type (e.g., Registration or Encounter)
- Number of Files
- Number of Records
- File Size
- Time Period
- Export ID
- Extract / Audit Run Time
- Date / Time processed in the Integration Engine (IE)
- QURE 1 Processing Time
- File Alignment Processing Time
- Import Engine Processing Time (formerly ETL processing time)
- NDW Load Date & Time
- QURE 2 Run Date & Time
- Extract File Name
- Post-IE File Name
- Audit File Name
- Notes specific to any portion of file processing

Appendix B. Final Summary Report

This report is still under development.

Appendix C. Load Exception Report

******* SAMPLE 'DATA LOAD EXCEPTION REPORT' *******

* DATA WAREHOUSE *
NPIRS Data Load Exception Report
Load Date: Jun 17, 2004

For more information please send an email to: ITSCDataWarehouse@ihs.gov
For error details, please visit: www.xxxxx.xxx

Export ID: 14
Export file name: 0001013040616190427.BDW.OUT
Static ASUFAC of exporting box: 000101
Beginning Date: Apr 01, 2001
Ending Date: Sep 30, 2001
Run Location: SELLS HOSP
Load Status: Exceptions detected
Total Number of Visits Received: 39456
Total Number of Patients Received: 0

ERROR DESCRIPTION	FIELD NAME	COUNT
******** REJECTED RECORDS ************		
PATIENT REGISTRATION NOT FOUND	n/a	8
******** TRANSFORMATIONS ************		
USING CREATE DATE AS LAST MOD	DATE_LAST_MOD_TS	10,000
INVALID-DATE	ELIG_START_DC	800
INVALID-DATE	REG_CREATE__DC	700
PSEUDO SSN NOT STORED	SSN-PSEUDO	1
FIELD SIZE TOO LARGE FOR DATABASE	XXXXXX_XXXX	1
******** INFORMATIONAL **************		
REGISTRATION WITH NO CHARTS	n/a	10,000

Appendix D. QURE 1 Compare Elements List

Unlike QURE 2, the QURE 1 Elements list is a subset of NDW fields. Due to resource restraints, only these specific data fields were selected based on criticality of IHS data needs.

Registration Compare Fields

- Unique Registration Record ID
- Birth Date
- Death Date
- Cause of Death
- Gender
- SSN
- Creation Date
- Patient Name
- Street Address
- Address City
- State
- Zip Code
- Zip Code Extension
- Community of Residence
- Community of Residence Start Date
- Service Eligibility
- Veteran Flag
- Benefit Class
- Tribe Code
- Blood Quantum
- Registration Status Code
- Chart Facility Code
- Chart Number
- Chart Status
- Insurance Category
- Coverage Type Code
- Eligibility Start Date
- Policy Number
- State Code
- Plan Name
- Insurer Name
- Eligibility End Date

- Insurer EIN
- Policy Prefix / Suffix
- Relation to Insured

Encounter Compare Fields

- Unique Encounter Code
- Date Last Modified Timestamp
- Data Entry Creation Date
- Encounter Delete Flag
- Unique Registration Code
- Chart Registering Facility Code
- Chart Number at Registering Facility
- Visit Admission Date
- Location of Encounter
- Type
- Service Category
- Clinic
- Urine Protein Test
- Urine Protein Value
- Discharge Date
- Provider Sequence Number
- Provider Affiliation Code
- Provider Discipline Code
- Provider Type
- Procedure Sequence Number
- ICD9 Procedure Code
- Procedure Date
- HCPCS Code
- CPT Code for ICD9 Procedure
- Procedure Sequence Number
- HCPCS Code
- CPT Quantity
- ADA Sequence Number
- ADA Code
- ADA Units
- ADA Code Fee
- Diagnosis Sequence Number
- Diagnosis
- Cause of DX
- Cause of Injury
- Place of Injury